



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 07 DEC 2004

WIPO

PCT

PCT/1804/S2529

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03104444.9

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

DEN HAAG, DEN
THE HAGUE,
LA HAYE, LE

16/12/03



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

**Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation**

Anmeldung Nr.:
Application no.: 03104444.9
Demande n°:

Anmeldetag:
Date of filing: 28/11/03
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
Verfahren und Einrichtung zum Transkribieren eines Audiosignals

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:
State:
Pays:

Tag:
Date:
Date:

Aktenzeichen:
File no.
Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:

/

Am Anmeldetag benannte Vertragsstaaten:
Contracting states designated at date of filing:
Etats contractants désignés lors du dépôt:

AT/BG/BE/CH/CY/CZ/DE/DK/EE/ES/FI/FR/GB/GR/HU/IE/IT/LI/LU/MC/

Bemerkungen:
Remarks:
Remarques:

Verfahren und Einrichtung zum Transkribieren eines Audiosignals

Die Erfindung bezieht sich auf ein Verfahren zum Transkribieren eines
5 Audiosignals enthaltend Signalabschnitte in Text enthaltend Textabschnitte für ein
Dokument, welches Dokument zum Wiedergeben von Informationen vorgesehen ist,
welche Informationen zumindest teilweise zu den durch die Transkription erhaltenen
Textabschnitten korrespondieren.

Die Erfindung bezieht sich weiters auf eine Einrichtung zum Transkribieren
10 eines Audiosignal enthaltend Signalabschnitte in Text enthaltend Textabschnitte für ein
Dokument, welches Dokument zum Wiedergeben von Informationen vorgesehen ist,
welche Informationen zumindest teilweise zu den durch die Transkription erhaltenen
Textabschnitten korrespondieren.

Die Erfindung bezieht sich weiters auf ein Computerprogrammprodukt, das
15 zum Transkribieren eines Audiosignals geeignet ist.

Die Erfindung bezieht sich weiters auf einen Computer, der das
Computerprogrammprodukt gemäß dem vorstehenden Absatz abarbeitet.

20 Ein solches Verfahren und eine solche Einrichtung und ein solches
Computerprogrammprodukt und ein solcher Computer sind aus dem Patentdokument
US 5,031,113 bekannt.

Bei der bekannten Einrichtung, mit deren Hilfe das bekannte Verfahren
durchführbar ist und die mit Hilfe des bekannten Computers realisiert ist, der das bekannte
25 Computerprogrammprodukt abarbeitet, wird auf Grundlage eines Audiosignals ein
Dokument erzeugt. Dabei werden in dem Audiosignal enthaltene Signalabschnitte als
Textabschnitte erkannt und gespeichert. Weiters werden Beziehungsdaten erzeugt und
gespeichert, die eine zeitliche Beziehung der Signalabschnitte mit den erkannten
Textabschnitten repräsentieren. Mit Hilfe der Einrichtung ist das Audiosignal über einen
30 Lautsprecher auf akustische Weise und das Dokument über einen Bildschirm auf visuelle
Weise wiedergebar. Die Beziehungsdaten werden bei einem akustischen Wiedergeben des
Audiosignals zum zeitsynchronen visuellen Hervorheben der zu den jeweiligen

Signalabschnitten zeitlich in Beziehung stehenden Textabschnitte verwendet, was in Fachkreisen unter dem Begriff „Synchrones Playback“ bekannt ist.

Bei der bekannten Einrichtung besteht das Problem, dass bei einem Dokument, das nicht nur den durch Transkription erzeugten Text, sondern auch andere Elemente, wie beispielsweise unveränderbare Formularfeldbezeichnungen oder Bilder oder Textbausteine oder audiovisuelle Objekte enthält, bei Anwendung des „Synchronen Playback“, und zwar insbesondere im Zusammenhang mit einem Kontrolllesen des durch Transkription erzeugten Textes durch eine Arbeitskraft, die den Text nicht selbst diktiert hat, erhebliche Schwierigkeiten auftreten, weil diese anderen nicht durch Transkription erzeugten Elemente nicht oder nur unzureichend berücksichtigbar sind.

Die Erfindung hat sich zur Aufgabe gestellt die Probleme bei einem Verfahren der eingangs im ersten Absatz angeführten Gattung und bei einer Einrichtung der eingangs im zweiten Absatz angeführten Gattung und bei einem Computerprogrammprodukt der eingangs im dritten Absatz angeführten Gattung und bei einem Computer der eingangs im vierten Absatz angeführten Gattung zu beseitigen und ein verbessertes Verfahren und eine verbesserte Einrichtung und ein verbessertes Computerprogrammprodukt und einen verbesserten Computer zu schaffen.

Zur Lösung der vorstehend angeführten Aufgabe sind bei einem Verfahren gemäß der Erfindung erfindungsgemäße Merkmale vorgesehen, so dass ein Verfahren gemäß der Erfindung auf die nachfolgend angegebene Weise charakterisierbar ist.

Verfahren zum Transkribieren eines Audiosignals enthaltend Signalabschnitte in Text enthaltend Textabschnitte für ein Dokument, welches Dokument zum Wiedergeben von Informationen vorgesehen ist, welche Informationen zumindest teilweise zu den durch die Transkription erhaltenen Textabschnitten korrespondieren, welches Verfahren die nachfolgend angeführten Schritte aufweist, nämlich:

Transkribieren der Signalabschnitte in Textabschnitte und Erzeugen von Beziehungsdaten, die zumindest eine zeitliche Beziehung zwischen je mindestens einem Signalabschnitt und je mindestens einem durch die Transkription erhaltenen Textabschnitt repräsentieren, und Erkennen einer Struktur des Dokuments und Abbilden der erkannten Struktur des Dokuments in den Beziehungsdaten.

Zur Lösung der vorstehend angeführten Aufgabe sind bei einer Einrichtung gemäß der Erfindung erfindungsgemäße Merkmale vorgesehen, so dass eine Einrichtung gemäß der Einrichtung auf die nachfolgend angeführte Weise charakterisierbar ist:

- Einrichtung zum Transkribieren eines Audiosignal enthaltend Signalabschnitte
- 5 in Text enthaltend Textabschnitte für ein Dokument, welches Dokument zum Wiedergeben von Informationen vorgesehen ist, welche Informationen zumindest teilweise zu den durch die Transkription erhaltenen Textabschnitten korrespondieren, mit Transkriptionsmitteln zum Transkribieren der Signalabschnitte in Textabschnitte und mit Beziehungsdaten-Erzeugungsmitteln, die zum Erzeugen von Beziehungsdaten ausgebildet sind, welche
- 10 Beziehungsdaten zumindest eine zeitliche Beziehung zwischen je mindestens einem Signalabschnitt und je mindestens einem durch die Transkription erhaltenen Textabschnitt repräsentiert, und mit Struktur-Erkennungsmitteln, die zum Erkennen einer Struktur des Dokuments ausgebildet sind, und mit Struktur-Abbildungsmitteln, die zum Abbilden der erkannten Struktur des Dokuments in den Beziehungsdaten ausgebildet sind.

- 15 Zur Lösung der vorstehend angeführten Aufgabe ist bei einem Computerprogrammprodukt, das zum Transkribieren eines Audiosignals geeignet ist, gemäß der Erfindung vorgesehen, dass das Computerprogrammprodukt direkt in einen Speicher eines Computers geladen werden kann und Softwarecodeabschnitte umfasst, wobei mit dem Computer das Verfahren gemäß der Erfindung abgearbeitet werden kann,
- 20 wenn das Computerprogrammprodukt auf dem Computer abgearbeitet wird.

Zur Lösung der vorstehend angeführten Aufgabe ist bei einem Computer gemäß der Erfindung vorgesehen, dass der Computer eine Recheneinheit und einen internen Speicher aufweist, und das Computerprogrammprodukt gemäß dem vorstehend angeführten Absatz abarbeitet.

- 25 Durch das Vorsehen der Maßnahmen gemäß der Erfindung ist der Vorteil erhalten, dass eine Struktur des zu erzeugenden Dokuments nicht nur in dem Dokument selbst manifestiert ist, sondern auch in den Beziehungsdaten manifestiert ist, wodurch wesentlich komplexere Dokumente erzeugbar und vor allem auch auf audiovisuelle Weise weiterverarbeitbar sind.

- 30 Durch das Vorsehen der zusätzlichen Maßnahmen gemäß dem Anspruch 2 bzw. dem Anspruch 9 ist weiters der Vorteil erhalten, dass eine bereits existierende Struktur in einem als Vorlage vorbereiteten Dokument, wie beispielsweise eine durch

vordefinierte Formularfelder gegebene Struktur eines Dokuments, auf zuverlässige Weise in den Beziehungsdaten abgebildet wird.

Durch das Vorsehen der zusätzlichen Maßnahmen gemäß dem Anspruch 3 bzw. dem Anspruch 10 ist weiters der Vorteil erhalten, dass die Struktur eines Dokuments,
5 die nur durch Strukturanweisungen, die in dem zu transkribierenden Audiosignal enthalten sind, weil sie beispielsweise von einer Person diktiert wurden, praktisch in Echtzeit, also während des Transkribierens erkannt und auf zuverlässige Weise in den Beziehungsdaten abgebildet wird.

Bei einer erfindungsgemäßen Lösung kann beispielsweise vorgesehen sein,
10 dass für jedes erkannte Strukturelement des Dokuments eine eigene Datei mit Beziehungsdaten erzeugt wird, also ein physikalisches Gruppieren der Beziehungsdaten erfolgt. Als besonders vorteilhaft hat es sich jedoch erwiesen, wenn zusätzlich die Maßnahmen gemäß dem Anspruch 4 bzw. dem Anspruch 11 vorgesehen sind, weil damit ein möglichst einfaches und zuverlässiges Gruppieren in einer einzigen Datei realisierbar
15 ist, so dass ein relativ aufwändiges Bearbeiten von mehreren Dateien vermieden ist. In diesem Fall kann das Gruppieren der Beziehungsdaten beispielsweise durch ein Kennzeichnen der Beziehungsdaten mit Hilfe von Strukturdaten erfolgen, die die erkannte Struktur des Dokuments repräsentieren. Es kann jedoch auch vorgesehen sein, dass die strukturell zusammengehörenden Beziehungsdaten abschnittsweise in der einzigen Datei
20 gruppiert werden, wobei jeder Abschnitt einem Strukturelement der erkannten Struktur des Dokuments zugeordnet ist.

Durch das Vorsehen der Maßnahmen gemäß dem Anspruch 5 bzw. dem Anspruch 12 ist weiters der Vorteil erhalten, dass die Effizienz bei dem Erkennen von Textabschnitten erhöht wird. Dies insbesondere deshalb, weil beispielsweise im Fall eines
25 Dokuments, das einen Bericht eines Radiologen repräsentiert, bei einem Transkribieren von administrativen Anweisungen des Radiologen nicht der radiologische Kontext benötigt wird, sondern ein wesentlich eingeschränkterer allgemeine Anweisungen betreffender Kontext ausreicht. Gleiches gilt für den Fall, dass eine Zusammenfassung eines Berichts transkribiert werden soll und beispielsweise im Wesentlichen im vorhinein bekannt ist,
30 dass in der Zusammenfassung hauptsächlich Standardformulierungen oder Standardphrasen zur Anwendung kommen. Gleiches gilt für den Fall, dass die Struktur in einem Dokument durch unterschiedliche Sprachen gegeben ist, die beispielsweise

abschnittsweise zur Anwendung kommen. So ist beispielsweise bei einer Verfügbarkeit eines ersten Sprachmodells oder eines zweiten Sprachmodells sichergestellt, dass das Transkribieren unter automatischer Auswahl des jeweiligen Sprachmodells erfolgt und gegebenenfalls das Dokument nachfolgend gemäß der Struktur, die durch die beiden
5 unterschiedlichen Sprachen gegeben ist, selektiv von unterschiedlichem Korrekturpersonal weiterbearbeitet wird.

Durch das Vorsehen der Maßnahmen gemäß dem Anspruch 6 bzw. dem Anspruch 13 ist der Vorteil erhalten, dass alle durch Transkription entstandenen textuellen Elemente des Dokuments problemlos und vor allem in der richtigen Reihenfolge
10 zusammenhängend wiedergebar sind, wobei nicht-textuale Elemente ausgelassen werden.

Durch das Vorsehen der Maßnahmen gemäß dem Anspruch 7 bzw. dem Anspruch 14 ist der Vorteil erhalten, dass ein zusammenhängendes akustisches Wiedergeben von Textabschnitten durchführbar ist, die einerseits durch das Transkribieren des Audiosignal erzeugt wurden und die andererseits auf andere Weise als durch das
15 Transkribieren des Audiosignals entstanden sind. Solche auf andere Weise entstandenen Textabschnitte können beispielsweise durch ein manuelles Eingeben von Text in das Dokument oder durch ein Einfügen von vordefinierten Textelementen bzw. Textobjekten, wie beispielsweise Feldbezeichnungen eines Formulars, oder durch ein Einfügen von vordefinierten Textbausteinen oder durch ein Korrigieren des durch Transkription
20 entstandenen Textes entstanden sein.

Die vorstehend angeführten Aspekte und weitere Aspekte der Erfindung gehen aus dem nachfolgend beschriebenen Ausführungsbeispiel hervor und sind anhand dieses Ausführungsbeispiels erläutert.

25

Die Erfindung wird im Folgenden anhand von einem in den Zeichnungen dargestellten Ausführungsbeispiel weiter beschrieben, auf das die Erfindung aber nicht beschränkt ist.

Die Figur 1 zeigt auf schematische Weise in Form eines Blockschaltbildes eine
30 Einrichtung gemäß einem Ausführungsbeispiel der Erfindung.

Die Figur 2 zeigt im Klartext eine Information, die in einem mit Hilfe der Einrichtung gemäß Figur 1 bearbeiteten Dokument enthalten ist.

Die Figur 3 zeigt im Klartext hinsichtlich einer Struktur des Dokuments gemäß der Figur 2 gegliederte Beziehungsdaten, die zumindest eine zeitliche Beziehung zwischen Signalabschnitten eines Audiosignals und Textabschnitten eines Text des Dokuments wiedergeben.

5

In der Figur 1 ist eine Einrichtung 1 dargestellt, die zum Transkribieren eines Audiosignals AS enthaltend Signalabschnitte SP in Text enthaltend Textabschnitte TP für ein Dokument DO ausgebildet. Das Audiosignal repräsentiert ein Diktat eines Sprechers. In der Figur 2 ist ein Dokument DO dargestellt, das zum Wiedergeben von Informationen vorgesehen ist, welche Informationen zumindest teilweise zu den durch die Transkription erhaltenen Textabschnitte TP korrespondieren. Im vorliegenden Fall weist das Dokument DO zu den transkribierten Textabschnitten TP nicht korrespondierende Template-Abschnitte auf, wie beispielsweise vordefinierte Formularfeldbezeichnungen „Author:“ oder „Date:“, die in einer Dokumentenvorlage fix vorgegebenen sind.

Die Einrichtung 1 weist einen ersten Eingang IN1 auf, an dem ihr das Audiosignal AS zuführbare ist. Es sei erwähnt, dass das Audiosignal AS auch auf andere Weise, wie beispielsweise mit Hilfe eines Datenträgers oder über ein Datennetzwerk in Form einer digitalen Repräsentation zuführbar sein kann, wenn die Einrichtung 1 über in an sich bekannter Weise eingerichtete Mittel verfügt.

Die Einrichtung 1 weist weiters einen zweiten Eingang IN2 auf, an dem ihr Bearbeitungssignale WS zuführbar sind, worauf später noch im Detail eingegangen ist.

Die Einrichtung 1 weist weiters Transkriptionsmittel 2 auf, die zum Empfangen des Audiosignals AS und zum Transkribieren der Signalabschnitte SP in die Textabschnitte TP ausgebildet sind. In diesem Zusammenhang sei erwähnt, dass es für den Fachmann eine Selbstverständlichkeit ist, das Audiosignal AS entsprechend zu konditionieren, wobei beispielsweise Filtermittel und Konvertierungsmittel zum Konvertieren in eine digitale Repräsentation zur Anwendung kommen, worauf nicht näher eingegangen wird. Das Transkribieren der Signalabschnitte SP erfolgt unter Berücksichtigung von in der Figur 1 nicht explizit dargestellten Sprecherdaten und einem selektierbaren Kontext. Kontextdaten, die ebenfalls in der Figur 1 nicht explizit dargestellt sind, repräsentieren die unterschiedlichen zur Auswahl stehenden Kontexte, wobei jeder Kontext eine Sprache, ein

Sprachemodell und ein Lexikon definiert bzw. umfasst. Die Sprecherdaten sind für den jeweiligen Sprecher repräsentativ. Auf Grundlage des zugeführten Audiosignals AS sind die Transkriptionsmittel 2 zum Erzeugen von Textdaten TXD ausgebildet, welche die erkannten Textabschnitte TP repräsentieren.

- 5 Die Einrichtung 1 weist weiters Dokumentdaten-Speichermittel 3 auf, die zum Speichern des Dokuments DO und der für das Dokument DO bestimmten Template-Daten TD und der Textdaten TXD ausgebildet und vorgesehen sind. Die Transkriptionsmittel 2 sind zum Zusammenwirken mit den Dokumentdaten-Speichermitteln 3 ausgebildet, so dass die Textdaten TXD in die dafür vorgesehenen Bereiche des Dokuments DO eingefügt
10 werden können. Mit Hilfe der Dokumentdaten-Speichermittel 3 sind weiters Objektdaten OD speicherbar, welche in das Dokument DO eingefügte Objekte OO repräsentieren, worauf nachfolgend noch eingegangen ist.

- Die Einrichtung 1 weist weiters Dokument-Bearbeitungsmittel 4 auf, die zum Empfangen von Bearbeitungssignalen WS über den zweiten Eingang IN2 ausgebildet sind.
15 Die Dokument-Bearbeitungsmittel 4 sind weiters unter Berücksichtigung des Bearbeitungssignals WS zum Erzeugen und zum Abgeben von Bearbeitungsdaten WD ausgebildet, die zum Verändern der mit Hilfe einer Transkription der Signalabschnitte SP erzeugten Textabschnitte TP in den Dokumentdaten-Speichermitteln 3 vorgesehen sind. Mit Hilfe der Dokument-Bearbeitungsmittel 4 sind beispielsweise die in der Figur 2
20 dargestellten und offensichtlich falsch erkannten Textabschnitte TP zwischen den Zeitmarken t93 und t100 korrigierbar, was durch eine Durchstreichung dieser Textabschnitte TP zwischen den Textmarken t93 und t100 und durch eine Einfügung von korrigierten Textabschnitten TP' zwischen der Textmarke t100 und t101 veranschaulicht ist. Für die durch Korrekturmaßnahmen erhaltenen weiteren Textabschnitte TP' gibt es
25 keine korrespondierenden Signalabschnitte SP in dem Audiosignal AS, da sie manuell eingefügt wurden. Gleiches gilt für das in der Figur 2 dargestellte Objekt OO.

- Die Transkriptionsmittel 2 sind weiters zum Erzeugen und zum Abgeben einer Information betreffend einen Anfangszeitpunkt t_n und einen Endzeitpunkt t_m eines Signalabschnitts SP innerhalb des Audiosignals AS und einer Information betreffend eine
30 Textabschnitt-Nummer WN ausgebildet, welche die Nummer des jeweils mit Hilfe der Transkriptionsmittel 2 erzeugten Textabschnitts TP repräsentiert.

Die Einrichtung 1 weist weiters Beziehungsdaten-Erzeugungsmittel 5 auf, die

zum Erzeugen von Beziehungsdaten RD ausgebildet sind, welche Beziehungsdaten RD eine zeitliche Beziehung zwischen je einem Signalabschnitt SP und je mindestens einem transkribierten Textabschnitt TP repräsentieren. Zu diesem Zweck sind die Beziehungsdaten-Erzeugungsmittel 5 zum Empfangen und zum Verarbeiten der

5 Information betreffend einen Anfangszeitpunkt t_n und einen Endzeitpunkt t_m der Signalabschnitte SP innerhalb des Audiosignals AS und der Information betreffend eine Textabschnitt-Nummer WN ausgebildet. Die Beziehungsdaten-Erzeugungsmittel 5 sind weiters zum Abgeben der Beziehungsdaten RD ausgebildet.

Die Einrichtung 1 weist weiters Struktur-Erkennungsmittel 6 auf, die zum

10 Erkennen einer Struktur des Dokuments DO ausgebildet sind, worauf nachfolgend im Detail eingegangen ist.

Zum Zweck des Erkennens der Struktur des Dokuments DO weisen die Struktur-Erkennungsmittel 6 eine erste Analysestufe 7 auf, die zum Analysieren des Dokuments DO hinsichtlich einer Struktur ausgebildet ist. Die erste Analysestufe 6 ist zum

15 Zugreifen auf die Dokumentdaten-Speichermittel 3 und zum Auslesen und Berücksichtigen der Template-Daten TD ausgebildet. Die erste Analysestufe 6 ist als ein Ergebnis ihrer Analyse zum Abgeben von ersten Analysedaten AD1 ausgebildet, welche eine auf Grundlage der Template-Daten TD erkennbare Struktur des Dokuments DO repräsentieren. Im vorliegenden Fall betrifft diese erkennbare Struktur das Vorliegen von zwei für eine

20 Eingabe von Text vorgesehenen Formularfeldern, die benachbart zu den beiden Formularfeldbezeichnungen „Author:“ und „Date“ angeordnet sind. Die erkennbare Struktur kann jedoch auch durch Bilder oder unveränderbare Textstellen gegeben sein. Es sei an dieser Stelle erwähnt, dass abgesehen von für einen Benutzer des Dokuments sichtbaren Strukturelementen auch im Normalgebrauch des Dokuments unsichtbare

25 Strukturelemente berücksichtigt werden, die durch Einstellungen definiert sind, die beispielsweise bei gängigen Textverarbeitungsprogrammen als sogenannte Bookmarks oder sogenannte Gliederungen bekannt sind und die nicht zu der durch das Dokument für den Benutzer wiedergebbaren Information gezählt werden können, weil sie hauptsächlich im Zusammenhang mit einem Steuern von Eingaben, einem Steuern von Ausgaben oder

30 einem Automatisieren der Verarbeitung des Dokuments zum Tragen kommen.

Zum Zweck des Erkennens der Struktur des Dokuments DO weisen die Struktur-Erkennungsmittel 5 weiters eine zweite Analysestufe 8 auf, die zum Analysieren

der erhaltenen Textabschnitte TP hinsichtlich einer Struktur des Dokuments DO ausgebildet ist. Die zweite Analysestufe 8 ist zum Empfangen der aus den Signalabschnitten SP transkribierten Textdaten TXD und zum Analysieren der Textdaten TXD im Hinblick auf durch den Sprecher gesprochene Strukturanweisungen ausgebildet, wobei die Strukturanweisungen zum Erzeugen und/oder Verändern und/oder Festlegen einer Struktur in dem Dokument DO vorgesehen bzw. geeignet sind. Dies kann beispielsweise gesprochene Formatzuweisungen, wie zum Beispiel ein Zuordnen von Überschriftsformaten, die zum Formatieren von Überschriften vorgesehen sind, zu einzelnen Textstellen, die als Überschriften formatiert werden sollen, oder auch ein durch gesprochene Befehle bewirktes Einfügen, Löschen oder Überschreiben von Textabschnitten TP betreffen.

Die zweite Analysestufe 8 ist weiters zum Empfangen der Bearbeitungsdaten WD und zum Analysieren der Bearbeitungsdaten WD im Hinblick auf eine mit Hilfe der Bearbeitungsdaten WD verursachte Veränderung einer bereits bestehenden Struktur des Dokuments DO oder im Hinblick auf eine neu definierte Struktur in dem Dokument DO ausgebildet. Dies kann beispielsweise ein Verändern einer Hierarchie von Überschriften oder ein Einfügen oder Entfernen von Elementen, wie beispielsweise Bildern, Texten oder Objekten betreffen, zu denen keine korrespondierenden Signalabschnitte SP in dem Audiosignal AS existieren. Es sei an dieser Stelle auch erwähnt, dass die zweite Analysestufe 8 auch zum Zugreifen auf die Dokumentdaten-Speichermittel 3 und zum Analysieren der durch Sprache oder manuelle Bearbeitung entstandenen Struktur des Dokuments DO ausgebildet sein kann.

Die zweite Analysestufe 8 ist in Analogie zu der ersten Analysestufe 7 zum Abgeben von das Ergebnis der Analyse repräsentierenden zweiten Analysedaten AD2 ausgebildet.

Die Einrichtung 1 weist weiters Struktur-Abbildungsmittel 9 auf, die zum Empfangen der ersten Analysedaten AD1 und der zweiten Analysedaten AD2 und der Beziehungsdaten RD ausgebildet sind. Die Struktur-Abbildungsmittel 9 sind unter Zuhilfenahme der ersten Analysedaten AD1 und der zweiten Analysedaten AD2 zum Abbilden der durch die Analysedaten AD1 und AD2 repräsentierten bzw. erkannten Struktur des Dokuments DO in den Beziehungsdaten RD ausgebildet. Die Struktur-Abbildungsmittel 9 sind weiters zum Abgeben von hinsichtlich der Struktur des

Dokuments DO strukturierten Beziehungsdaten SRD ausgebildet, die im vorliegenden Fall eine in der Figur 3 dargestellte logische Gruppierung der Beziehungsdaten RD darstellen.

Die Einrichtung 1 weist weiters Beziehungsdaten-Speichermittel 10 auf, die zum Speichern der strukturierten Beziehungsdaten SRD ausgebildet sind. Die Struktur-
5 Abbildungsmittel 9 sind zum Zugreifen auf die Beziehungsdaten-Speichermittel 10 vorgesehen, wobei die strukturierten Beziehungsdaten SRD in den Beziehungsdaten-Speichermitteln 10 speicherbar bzw. bereits gespeicherte Beziehungsdaten SRD veränderbar sind.

In der Figur 3 ist im Klartext eine Abbildung der strukturierten
10 Beziehungsdaten SRD für das in der Figur 2 dargestellte Dokument DO wiedergegeben. Die Figur 3 zeigt zeilenweise aufgelistete Einträge die zu Elementen des Dokuments DO korrespondieren und mit Hilfe der Zahlen Eins (1) bis Sechsfundfünfzig (56) nummeriert sind. Eine erste Spalte C1 gibt die Nummer des jeweiligen Dokumenteintrags wieder. Eine zweite Zeile C2 gibt den jeweiligen Anfangszeitpunkt eines Signalsabschnitts SP innerhalb
15 des Audiosignals AS wieder, der zu dem durch die jeweilige Nummer korrespondierenden Element des Dokuments DO, wie beispielsweise einem aus einem Signalabschnitt SP transkribierten Textabschnitt TP, korrespondiert. Eine dritte Spalte C3 gibt den jeweiligen Endzeitpunkt des zuvor erwähnten Signalabschnitts SP innerhalb des Audiosignals AS wieder. Wie aus der Figur 3 ersichtlich ist, betreffen die mit Hilfe der strukturierten
20 Beziehungsdaten repräsentierten Dokumenteinträge jedoch nicht nur solche Elemente, die mit Hilfe der Transkription des Audiosignals AS erzeugt wurden, sondern auch solche Elemente, die auf andere Art erzeugt wurden und die in dem Dokument zwischen den Signalabschnitten SP des Audiosignals AS lokalisiert sind, wie beispielsweise die Elemente der Zeile 40 und 52. Eine Spalte C4 repräsentiert für den jeweiligen
25 Dokumenteintrag die Zugehörigkeit zu einer in dem Dokument DO enthaltenen Struktur. Dabei ist insbesondere darauf hingewiesen, dass auch Dokumenteinträge, wie beispielsweise solche zwischen den Zeitmarken t78 und t79 bzw. zwischen den Zeitmarken t100 und t101 registrierten Dokumenteinträgen, in den Beziehungsdaten RD manifestiert sind, für welche Dokumenteinträge kein Audiosignal AS existiert, um später
30 gegebenenfalls eine solche Elemente umfassende oder eine solche Elemente auslassende Audiowiedergabe des Audiosignals AS oder eine Nachvollziehbarkeit der Entstehung und/oder Veränderung des Dokuments gewährleisten zu können.

Die Einrichtung 1 weist weiters Audiodaten-Speichermittel 11 auf, die zum Speichern von Audiodaten AD ausgebildet sind, welche das Audiosignal AS repräsentieren und von den Transkriptionsmitteln 2 an die Audiosignal-Speichermittel 11 abgegeben werden. Die Audiodaten AD repräsentieren das Audiosignal AS auf an sich bekannte
5 Weise in einer digitalen Darstellung, in der auf die Signalabschnitte SP für eine spätere Wiedergabe des Audiosignals AS unter Berücksichtigung der strukturierten Beziehungsdaten SRD zurückgegriffen werden kann.

Die Transkriptionsmittel 2 sind weiters in Abhängigkeit von der erkannten Struktur des Dokuments DO, also in Abhängigkeit von den strukturierten Beziehungsdaten
10 SRD konfigurierbar, wobei im vorliegenden Fall zwischen drei verschiedenen Kontexten in Abhängigkeit von der Struktur ausgewählt wird. So wird bei einem Erkennen, dass es sich um ein Strukturelement „report heading“ handelt, ein erster Kontext ausgewählt, und dass es sich um ein Strukturelement „chapter heading“ handelt, ein zweiter Kontext ausgewählt, und dass es sich um ein Strukturelement „text“ handelt der dritte Kontext
15 ausgewählt. Dadurch ist sichergestellt, dass, so bald das Strukturelement „text“ vorliegt, der Kontext mit dem maximalen Lexikonumfang bereitgestellt wird, was üblicherweise für das Transkribieren von Signalabschnitten SP, die das Strukturelement „report heading“ oder „chapter heading“ betreffen, nicht notwendig ist. Weiters wird bei einem Erkennen, dass es sich um das Strukturelement „author“ handelt, ein im wesentlichen Namen
20 betreffender vierter Kontext ausgewählt. Weiters wird bei einem Erkennen, dass es sich um das Strukturelement „date“ handelt ein im wesentlichen Datumsangaben betreffender fünfter Kontext ausgewählt.

Es sei an dieser Stelle erwähnt, dass unter Berücksichtigung der erkannten Struktur auch die Sprache oder das Sprachmodell oder auch eine Auswahl zwischen
25 verschiedenen Sprecherdaten erfolgen kann. Weiters sei erwähnt, dass das Berücksichtigen einer Struktur des Dokuments DO bei den Transkriptionsmitteln 2 nicht erst dann erfolgen muss, wenn die erkannte Struktur bereits in den strukturierten Beziehungsdaten SRD Einzug gehalten hat, sondern dass das Berücksichtigen der Struktur bereits auf Grundlage der ersten Analysedaten AD1 und/oder der zweiten Analysedaten AD2 erfolgen kann,
30 sobald diese von den Strukturerkennungsmitteln 6 beispielsweise direkt an die Transkriptionsmittel 2 abgegeben werden.

Die Einrichtung 1 weist weiters Adaptierungsmittel 12 auf, die unter

Zuhilfenahme der strukturierten Beziehungsdaten SRD zum Adaptieren des jeweiligen Kontexts für die Transkriptionsmittel 2 ausgebildet sind. Zu diesem Zweck sind die Adaptierungsmittel 12 zum Auslesen der strukturierten Beziehungsdaten SRD aus den Beziehungsdaten-Speichermitteln 9 und zum Auslesen der Textdaten TXD aus den

5 Dokument-Speichermitteln 3 und zum Analysieren der Textdaten TXD unter Heranziehung der strukturierten Beziehungsdaten SRD und/oder zum Analysieren der nach dem erstmaligen Erzeugen und Speichern der Textdaten TXD mit Hilfe der strukturierten Beziehungsdaten SRD protokollierten Veränderungen der Textdaten TXD ausgebildet. Als ein Ergebnis des Analysierens der Textdaten TXD sind die Adaptierungsmittel 12 zum

10 Abgeben einer Veränderung- bzw. Adaptierungsinformation CI an die Transkriptionsmittel 2 ausgebildet, mit deren Hilfe der jeweilige Kontext adaptierbar ist, so dass in Zukunft bessere Ergebnisse bei dem Transkribieren erhalten werden.

Die Einrichtung 1 weist weiters Wiedergabe-Steuermittel 13 auf, die unter Berücksichtigung der erkannten Struktur des Dokuments DO zum Bewirken eines

15 akustischen Wiedergebens der Signalabschnitte SP des Audiosignals AS zeitsynchron zu einem visuellen Hervorheben der transkribierten Textabschnitte TP bei einem visuellen Wiedergeben der Textabschnitte TP des Dokuments DO ausgebildet sind. Zu diesem Zweck sind die Wiedergabe-Steuermittel 13 zum Zugreifen auf die in den Beziehungsdaten-Speichermitteln 10 gespeicherten strukturierten Beziehungsdaten SRD

20 und zum Zugreifen auf die in den Dokument-Speichermitteln 3 gespeicherten Textdaten TXD ausgebildet, welche mit Hilfe der strukturierten Beziehungsdaten SRD als solche Textdaten TXD gekennzeichnet sind, für die Signalabschnitte SP existieren, die mit Hilfe der Audiodaten AD repräsentiert sind. Die Wiedergabe-Steuermittel 13 sind weiters zum Zugreifen auf die Signalabschnitte SP in den Audiodaten AD ausgebildet, welche

25 Signalabschnitte SP durch die jeweiligen in den strukturierten Beziehungsdaten SRD protokollierten Zeitmarken t_n und t_m zeitlich begrenzt sind. Die Wiedergabe-Steuermittel 13 sind weiters zum zeitsynchronen Abgeben von den die jeweiligen Signalabschnitte SP repräsentierenden Audiodaten AD an eine erste Wiedergabeeinrichtung 14 und zum Abgeben der zeitlich dazu korrespondierenden Textanzeige-Steuerdaten TDCD an eine

30 zweite Wiedergabeeinrichtung 15 ausgebildet. Mit Hilfe der Textanzeige-Steuerdaten TDCD ist erstens die Information des Dokuments DO an die zweite Wiedergabeeinrichtung 15 abgebar, die zum visuellen Wiedergeben dieser Information

ausgebildet ist, und zweitens ein synchrones Hervorheben des jeweiligen Textabschnitts TP definierbar, während der dazu korrespondierende Signalabschnitt SP in Form der Audiodaten AD an die erste Wiedergabeeinrichtung 14 abgegeben wird.

Im vorliegenden Fall ist sowohl die erste Wiedergabeeinrichtung 14, die durch
5 einen Audioverstärker mit integriertem Lautsprecher realisiert ist, als auch die zweite Wiedergabeeinrichtung 15, die durch einen Bildschirm realisiert ist, mit der Einrichtung 2 über jeweils einen zugeordneten Signalausgang OUT1 und OUT 2 verbunden. Es sei an dieser Stelle jedoch erwähnt, dass die beiden Einrichtungen 14 und 15 auch durch eine Kombinationseinrichtung gebildet sein können, die über einen einzigen Signalausgang der
10 Einrichtung 2 mit dieser verbunden sind. Weiters können die beiden Einrichtungen 14 und 15 auch in der Einrichtung 1 integriert sein.

Die Einrichtung 1 weist Sprach-Synthetisierungsmittel 16 auf, die zum Synthetisieren von Textdaten TXD in synthetische Sprache ausgebildet sind und die dazu dient, für jene Textabschnitte TP', für die keine Signalabschnitte SP in dem Audiosignal
15 AS existieren, auf synthetische Weise einer akustischen Wiedergabe zugänglich zu machen. Die Sprach-Synthetisierungsmittel 16 sind eingangsseitig mit den Wiedergabe-Steuermitteln 13 und ausgangsseitig mit dem Signalausgang OUT1 verbunden.

Die Wiedergabe-Steuermittel 13 sind weiters zum Kooperieren mit den Sprach-Synthetisierungsmitteln 16 ausgebildet und unter Zuhilfenahme der Sprach-
20 Synthetisierungsmittel 16 zum Bewirken eines akustischen Wiedergebens von zusätzlich zu den durch Transkription des Audiosignals AS erhaltenen Textabschnitten TP erzeugten weiteren Textabschnitten TP' ausgebildet, welche weiteren Textabschnitte TP' benachbart zu den durch die Transkription des Audiosignals AS erhaltenen Textabschnitten TP in dem Dokument DO existieren. Dabei ist gegebenenfalls ein Unterbrechen des Wiedergebens des
25 Audiosignals AS während des Wiedergebens der weiteren Textabschnitte TP' unter Kontrolle der Wiedergabe-Steuermittel 13 bewirkbar, wenn diese weiteren Textabschnitte TP' beispielsweise als Bestandteil des Objekts OO oder durch Korrektur in das Dokument DO Eingang gefunden haben, wie dies an Hand der Figur 2 dargestellt ist.

Im Folgenden ist nun anhand eines Anwendungsbeispiels für die Einrichtung 1
30 gemäß der Figur 1 die Arbeitsweise der Einrichtung 1 erläutert.

Gemäß dem Anwendungsbeispiel sei angenommen, dass ein Geschäftsmann einen Bericht betreffend einen Geschäftsplan diktiert. Mit Hilfe eines an dem ersten

Eingang IN1 angeschlossenen Mikrofons 17 wird das Audiosignal AS erzeugt und der Einrichtung 1 zugeführt.

Mit Hilfe der Einrichtung 1 ist ein Verfahren zum Transkribieren des Audiosignals AS durchführbar. Dabei ist zu Beginn des Diktats das in der Figur 2 in
5 seinem finalen Bearbeitungszustand abgebildete Dokument DO im wesentlichen leer und weist lediglich die vordefinierten und unveränderbaren Template-Daten TD auf, die vordefinierte Formularfeldbezeichnungen, und zwar im vorliegenden Fall die Formularfeldbezeichnungen „Autor:“ und „Date:“, repräsentieren.

Bei dem Verfahren werden Signalabschnitte SP in dazu korrespondierende
10 Textabschnitte TP transkribiert und Beziehungsdaten RD erzeugt, welche die zeitliche Beziehung zwischen je einem Signalabschnitt SP und je mindestens einem transkribierten Textabschnitt TP repräsentieren.

Im vorliegenden Fall diktiert der Geschäftsmann zunächst die Wörter: „Autor: Michael Schneider“.

15 Um den Erkennungs- bzw. Transkription-Prozess zu verbessern, erfolgt mit Hilfe der Einrichtung 1 auch ein Erkennen einer Struktur des Dokuments DO und ein Abbilden der erkannten Struktur des Dokuments DO in den Beziehungsdaten RD. Zu diesem Zweck wird beginnend mit dem Empfangen des Audiosignals AS die Struktur des Dokuments DO mit Hilfe der ersten Analysestufe 7 analysiert und festgestellt, dass die
20 beiden zuvor erwähnten Formularfeldbezeichnungen existieren. Die ersten Analysedaten AD1 repräsentieren dieses Analyseergebnis, das mit Hilfe der Strukturabbildungsmittel 9 in den Beziehungsdaten RD durch ein Erzeugen der strukturierten Beziehungsdaten SRD abgebildet wird, was bei den Transkriptionsmitteln 2 dazu ausgenutzt wird, die Signalabschnitte, welche die gesprochenen Wörter: „Autor:“ repräsentieren, zu verwerfen.
25 Weiters wird für das Transkribieren der vierte Kontext selektiert, bei dem nur einige bekannte Namen zur Auswahl stehen. Dies beschleunigt und verbessert das Transkribieren der zwischen den in der Figur 2 dargestellten Textzeitmarken t1 bis t4 enthaltenen Wörter. In analoger Weise erfolgt das Transkribieren der Datumsangabe, die mit Hilfe von mehreren Signalabschnitten SP repräsentiert ist unter Verwendung von dem fünften
30 Kontext. Dabei werden die zwischen den Zeitmarken t5 und t6 auftretenden Signalabschnitt SP zusammengefasst, weil die Transkriptionsmittel 2 bei einem Erkennen eines ein Datum anzeigenden Strukturelements ein vordefiniertes Datumsform anwenden.

Nach dem Diktieren der Einträge für die Formularfelder kann der Geschäftsmann eine beliebige Struktur für den nachfolgenden Text definieren. Um dem Rechnung zu tragen, erfolgt gemäß dem Verfahren ein Analysieren der erkannten Textabschnitte TP, also der Textdaten TXD, hinsichtlich der zu erzeugenden Struktur des Dokuments DO. So wird beispielsweise von dem Geschäftsmann der Satz :“Report heading Business Plan Report“ diktiert. Mit Hilfe der zweiten Analysestufe 8 wird daraufhin unter Heranziehung der erkannten Textabschnitte TP erkannt, dass es sich um ein Strukturelement betreffend die Hauptüberschrift des Dokuments DO handelt.

Dem gemäß werden die zwischen den Zeitmarken t7, t8 und t9, t10 und t11, t12 erkannten Textabschnitte TP dem Strukturelement „report heading“ zugeordnet, so wie dies in der Figur 3 dargestellt ist, wobei eine logische Gruppierung der Beziehungsdaten RD als strukturierte Beziehungsdaten SRD erfolgt.

Nachdem dieses Strukturelement anhand der Wörter „report heading“ erkannt wurde, erfolgt bei den Transkriptionsmitteln 2 auf Grundlage des erkannten Strukturelements ein Konfigurieren der Transkriptionsmittel 2 dahingehend, dass der zweite Kontext verwendet wird, der möglichst allgemeine Ausdrücke für Überschriften im Geschäftsalltag enthält.

Der Geschäftsmann setzt sein Diktat mit den Worten „chapter heading introduction“ fort, was ebenfalls dazu führt, dass ein weiteres Strukturelement, nämlich das Strukturelement „chapter heading“ erkannt wird. In diesem Fall wird der zweite Kontext ausgewählt, der jedoch im Vergleich zu dem Kontext betreffend die Hauptüberschrift einen größeren Lexikonumfang aufweist. Weiters wird der erkannte Textabschnitt TP, der zu dem Signalabschnitt SP zwischen den Zeitmarken t13 und t14 korrespondiert, in den Beziehungsdaten-Speichermitteln 9 durch das Strukturelement „chapter heading“ gekennzeichnet.

Da im nachfolgend gesprochenen Satz, der durch Signalabschnitte SP zwischen den Zeitmarken t15 bis t44 repräsentiert ist, keine weiteren gesprochenen Strukturanweisungen auftreten, wird für das Transkribieren der das größte Lexikon umfassende Kontext ausgewählt und die Beziehungsdaten RD für diese Signalabschnitte SP dem Strukturelement „text“ zugeordnet.

Danach wird wieder anhand des diktierten Textes das Strukturelement „chapter heading“ erkannt und der Textabschnitt TP, der zu dem Signalabschnitt zwischen den

Zeitmarken t45 und t46 korrespondiert, diesem Strukturelement logisch zugeordnet.

Der nachfolgend gesprochene Satz, der durch die Zeitmarken t47 bis t78 begrenzt ist, wird wegen eines Fehlens jeglicher erkennbarer Strukturelemente dem Strukturelement „text“ zugeordnet, wobei zum Transkribieren wiederum der das größte
5 Lexikon aufweisende dritte Kontext herangezogen wird.

Danach wird von dem Geschäftsmann ein Objekt OO in das Dokument DO eingefügt, das sowohl eine Grafik als auch einen Text aufweist, zu welchem Text jedoch kein Audiosignal AS korrespondiert, weil er durch eine textuale Eingabe erzeugt wurde. Das Einfügen des Objekts OO erfolgt im vorliegenden Fall unter Zuhilfenahme von
10 taktilen Eingabemitteln 18, nämlich einer Tastatur, die an dem zweiten Eingang IN2 angeschlossen ist, und der Textbearbeitungsmittel 4. Es sei jedoch erwähnt, dass das Einfügen des Objekts OO auch durch gesprochene Befehle, die mit Hilfe der Transkriptionsmittel 2 transkribiert und von weiteren in der Einrichtung 1 nicht dargestellten Mitteln als Befehle erkannt und ausgeführt werden, erzeugt werden kann.
15 Dem gemäß wird im vorliegenden Fall das Einfügen des Objekts OD mit Hilfe der zweiten Analysestufe 8 erkannt und in den Beziehungsdaten-Speichermitteln 9 das Vorliegen dieses Objekts zwischen den Zeitmarken t78 und t79 vermerkt.

Der nachfolgend diktierter Text zwischen den Zeitmarken t79 und t100 wird zunächst dem Strukturelement „text“ zugeordnet. Jedoch sind bei dem Transkribieren unter
20 Verwendung des dritten Kontexts Fehler zwischen den Zeitmarken t93 und t100 aufgetreten, die von dem Geschäftsmann mit Hilfe der Eingabemittel 18 korrigiert werden. Zu diesem Zweck werden die Textabschnitte TP zwischen den Zeitmarken t93 und t100 gelöscht und neue Textabschnitte TP' hinzugefügt, welche die gelöschten Textabschnitte TP ersetzen und vor der Zeitmarke t101 angesiedelt sind. Mit Hilfe der zweiten
25 Analysestufe 8 wird diese Veränderung in dem Dokument DO registriert bzw. erkannt und die zwischen den Zeitmarken t93 und t100 ursprünglich vorgelegenen Textabschnitte TP durch das Strukturelement „text to skip“ markiert, so dass bei einem akustischen Wiedergeben der gespeicherten Audiodaten AD diese Textabschnitte TP übersprungen werden. Weiters werden die weiteren Textabschnitte TP', die manuell vor der Zeitmarke
30 t101 eingegeben wurden, durch das Strukturelement „text inserted: no audio“ gekennzeichnet, was definiert, dass es sich um einen diktierten Text handelt, der jedoch nachfolgend korrigiert oder überarbeitet wurde, und dass zu den neu hinzugekommenen

Textabschnitten TP' keine korrespondierenden Signalabschnitte SP in den gespeicherten Audiodaten AD enthalten sind.

Die nachfolgend in dem Diktat auftretenden Signalabschnitte SP werden in den Beziehungsdaten-Speichermitteln 9 durch das Strukturelement „text“ gekennzeichnet, weil
5 mit Hilfe der Struktur-Erkennungsmittel 5 keine anderen Strukturelemente erkennbar und daher auch nicht zuordenbar sind.

Nachfolgend an ein Diktieren des Textes und gegebenenfalls ein Korrigieren des diktierten Textes kann der Geschäftsmann gemäß dem Verfahren einen Wiedergabenmodus aktivieren, mit dessen Hilfe ein genaues audiovisuelles Verfolgen des
10 transkribierten Audiosignals AS zeitsynchron zu einem visuellen Hervorheben der zu den jeweils durch die Zeitmarken tn bzw. tm angegebenen Signalabschnitte SP korrespondierenden Textabschnitte TP ermöglicht ist, wobei das zeitsynchrone audiovisuelle Wiedergeben der Textabschnitte TP und der Signalabschnitte SP unter Ausnutzung der strukturierten Beziehungsdaten SRD erfolgt. Dadurch ist erreicht, dass
15 beispielsweise nicht diktierter Elemente des Dokuments OD bei dem visuellen Hervorheben übersprungen bzw. ignoriert werden.

Gemäß dem Verfahren ist weiters gewährleistet, dass die zusätzlich zu den durch das Transkribieren des Audiosignals AS erhaltenen Textabschnitte TP erzeugten weiteren Textabschnitte TP' unter Zuhilfenahme von auf synthetische Weise, also unter
20 Zuhilfenahme der Sprach-Synthetisierungsmittel 16, erzeugbare Sprache wiedergegeben werden. Das Verfahren gewährleistet weiters, dass das Wiedergeben des Audiosignals AS während des Wiedergebens der weiteren Textabschnitte TP' gegebenenfalls unterbrochen wird, wenn die weiteren Textabschnitte zwischen durch Transkription erzeugten Textabschnitten TP eingebettet sind.

25 Dadurch ist erreicht, dass auch eine Korrektur oder eine Einfügung entsprechend ihrer Position in dem Dokument DO in der richtigen Reihenfolge bzw. im richtigen Zusammenhang mit den durch Transkription entstandenen Textabschnitten TP beim Wiedergeben berücksichtigt wird.

Im vorliegenden Fall ist die Einrichtung 1 durch einen in der Figur 1 nicht
30 dargestellten Computer mit einer Recheneinheit und einem internen Speicher realisiert, der ein Computerprogrammprodukt abarbeitet. Das Computerprogrammprodukt ist auf einem in der Figur 1 nicht dargestellten computerlesbaren Datenträger bzw. Medium gespeichert,

beispielsweise auf einer DVD oder einer CD oder einem nicht-flüchtigen Halbleiterspeicher. Das Computerprogrammprodukt kann von dem computerlesbaren Medium in den internen Speicher des Computers geladen werden, so dass mit Hilfe des Computers das erfindungsgemäße Verfahren zum Transkribieren von Signalabschnitten SP
5 in Textabschnitte TP durchgeführt wird, wenn das Computerprogrammprodukt auf dem Computer abgearbeitet wird.

Es sei an dieser Stelle erwähnt, dass die Einrichtung 1 auch durch mehrere Computer realisiert sein kann, die über ein Computernetzwerk verteilt sind und als Computersystem zusammenarbeiten, so dass einzelne Funktionen der Einrichtung 1
10 beispielsweise von einzelnen Computern übernommen werden können.

Es sei erwähnt, dass das zusammenhängende Wiedergeben der Textabschnitte TP und der weiteren Textabschnitte TP' auch dann gewährleistet ist, wenn die auf andere Weise erhaltenen weiteren Textabschnitte TP' am Anfang oder am Ende der durch Transkription erhaltenen Textabschnitte TP angesiedelt sind.

15 Es sei erwähnt, dass die strukturierten Beziehungsdaten SRD auch gesprochen oder manuell aktivierte Befehle umfassen können, wodurch ein weiterer Beitrag zu einer Nachvollziehbarkeit betreffend die Entstehung der durch das Dokument wiedergebbaren Information erhalten ist.

Es sei weiters erwähnt, dass die erfindungsgemäße Einrichtung auch in
20 privatem Gebrauch oder in medizinischem Gebrauch oder im sicherheitstechnischen Gebrauch einsetzbar ist, wobei diese Aufzählung nicht abschließend ist.

Hinsichtlich die Zuordnung zwischen Signalabschnitten SP und durch die Transkription erhaltenen Textabschnitten TP sei erwähnt, dass beispielsweise das gesprochene Wort „Today“ als ein zusammengehörender Signalabschnitt SP erkannt wird
25 und dass daraus mehrere Textabschnitte TP, nämlich „31. Nov. 2003“ durch Transkription erzeugt werden, so dass im vorliegenden Fall die Beziehungsdaten RD die zeitliche Beziehung zwischen einem einzigen Signalabschnitt SP und drei Textabschnitten TP wiedergeben. In diesem Zusammenhang sei weiters erwähnt, dass die Zuordnung zwischen Signalabschnitten SP und durch die Transkription erhaltenen Textabschnitten TP auch
30 derart gegeben sein kann, dass beispielsweise das gesprochene Datum „31. Nov. 2003“, das durch mindestens drei Signalabschnitte SP repräsentiert ist, nämlich jene die das Wort „31“ und „November“ und „2003“ repräsentieren, zu einem einzigen Textabschnitt TP,

beispielsweise „Heute“ oder „Morgen“ oder „Gestern“, durch Transkription zusammengefasst werden, so dass im vorliegenden Fall die Beziehungsdaten RD die zeitliche Beziehung zwischen drei Signalabschnitten SP und einem Textabschnitt TP wiedergeben.

Patentansprüche:

1. Verfahren zum Transkribieren eines Audiosignals (AS) enthaltend Signalabschnitte (SP) in Text enthaltend Textabschnitte (TP) für ein Dokument (DO), welches Dokument (DO) zum Wiedergeben von Informationen vorgesehen ist, welche
5 Informationen zumindest teilweise zu den durch die Transkription erhaltenen Textabschnitten (TP) korrespondieren, welches Verfahren die nachfolgend angeführten Schritte aufweist, nämlich:
Transkribieren der Signalabschnitte (SP) in Textabschnitte (TP) und Erzeugen von Beziehungsdaten (RD), die zumindest eine zeitliche Beziehung zwischen je
10 mindestens einem Signalabschnitt (SP) und je mindestens einem durch die Transkription erhaltenen Textabschnitt (TP) repräsentieren, und Erkennen einer Struktur des Dokuments (DO) und Abbilden der erkannten Struktur des Dokuments (DO) in den Beziehungsdaten (RD).
2. Verfahren nach Anspruch 1,
15 wobei das Erkennen der Struktur des Dokuments (DO) durch ein Analysieren des Dokuments (DO) erfolgt.
3. Verfahren nach Anspruch 1,
wobei das Erkennen der Struktur des Dokuments (DO) durch ein Analysieren der erkannten Textabschnitte (TP) erfolgt.
- 20 4. Verfahren nach Anspruch 1,
wobei das Abbilden der erkannten Struktur des Dokuments (DO) durch ein logisches Gruppieren der Beziehungsdaten (RD) erfolgt.
5. Verfahren nach Anspruch 1,
wobei ein Konfigurieren von zum Transkribieren von Textabschnitten (TP) vorgesehenen
25 Transkriptionsmitteln (2) in Abhängigkeit von der erkannten Struktur erfolgt.
6. Verfahren nach Anspruch 1,
wobei ein akustisches Wiedergeben der Signalabschnitte (SP) des Audiosignals (AS) zeitsynchron zu einem visuellen Hervorheben der transkribierten Textabschnitte (TP) bei einem visuellen Wiedergeben der Textabschnitte (TP) erfolgt und dabei die erkannte
30 Struktur des Dokuments (DO) berücksichtigt wird.
7. Verfahren nach Anspruch 3,
wobei zusätzlich zu den durch die Transkription des Audiosignals (AS) erhaltenen

Textabschnitten (TP) erzeugt weitere Textabschnitte (TP'), welche weiteren Textabschnitte (TP') benachbart zu den durch die Transkription des Audiosignals (AS) erhaltenen Textabschnitten (TP) in dem Dokument (DO) existieren, unter Zuhilfenahme von auf synthetische Weise erzeugbare Sprache wiedergegeben werden, und

- 5 wobei gegebenenfalls das Wiedergeben des Audiosignals (AS) während des Wiedergebens der weiteren Textabschnitte (TP') unterbrochen wird.

8. Einrichtung (1) zum Transkribieren eines Audiosignal (AS) enthaltend Signalabschnitte (SP) in Text enthaltend Textabschnitte (TP) für ein Dokument (DO), welches Dokument (DO) zum Wiedergeben von Informationen vorgesehen ist, welche
- 10 Informationen-zumindest teilweise zu den durch die Transkription erhaltenen Textabschnitten (TP) korrespondieren, mit Transkriptionsmitteln (2) zum Transkribieren der Signalabschnitte (SP) in Textabschnitte (TP) und mit Beziehungsdaten-Erzeugungsmitteln (5), die zum Erzeugen von Beziehungsdaten (RD)
- 15 ausgebildet sind, welche Beziehungsdaten (RD) zumindest eine zeitliche Beziehung zwischen je mindestens einem Signalabschnitt (SP) und je mindestens einem durch die Transkription erhaltenen Textabschnitt (TP) repräsentiert, und mit Struktur-Erkennungsmitteln (6), die zum Erkennen einer Struktur des Dokuments (DO) ausgebildet sind, und
- 20 mit Struktur-Abbildungsmitteln (9), die zum Abbilden der erkannten Struktur des Dokuments (DO) in den Beziehungsdaten (RD) ausgebildet sind.

9. Einrichtung (1) nach Anspruch 8, wobei die Struktur-Erkennungsmittel (6) mit Hilfe einer ersten Analysestufe (7) realisiert sind, die zum Analysieren des Dokuments (DO) hinsichtlich seiner Struktur ausgebildet ist.

- 25 10. Einrichtung (1) nach Anspruch 8, wobei die Struktur-Erkennungsmittel (6) mit Hilfe einer zweiten Analysestufe (8) realisiert sind, die zum Analysieren der erhaltenen Textabschnitte (TP) hinsichtlich einer Struktur des Dokuments (DO) ausgebildet ist.

11. Einrichtung (1) nach Anspruch 8, wobei die Struktur-Abbildungsmittel (9) zum logischen Gruppieren der Beziehungsdaten (RD) ausgebildet sind.
- 30

12. Einrichtung (1) nach Anspruch 8,

wobei die Transkriptionsmittel (2) in Abhängigkeit von der erkannten Struktur konfigurierbar sind.

13. Einrichtung (1) nach Anspruch 8,
wobei Wiedergabe-Steuermittel (13) vorgesehen sind, die unter Berücksichtigung der
5 erkannten Struktur des Dokuments (DO) zum Bewirken eines akustischen Wiedergebens der Signalabschnitte (SP) des Audiosignals (AS) zeitsynchron zu einem visuellen Hervorheben der transkribierten Textabschnitte (TP) bei einem visuellen Wiedergeben der Textabschnitte (TP) ausgebildet sind.

14. Einrichtung (1) nach Anspruch 13,
10 wobei Sprache-Synthetisierungsmittel (16) vorgesehen sind, die zum Synthetisieren von Textabschnitten (TP, TP') in Sprache ausgebildet sind, und wobei die Wiedergabe-Steuermittel (13) unter Zuhilfenahme der Sprache-Synthetisierungsmittel (16) zum Bewirken eines akustischen Wiedergebens von zusätzlich zu den durch Transkription des Audiosignals erhaltenen Textabschnitten (TP) erzeugten
15 weiteren Textabschnitten (TP'), welche weiteren Textabschnitte (TP') benachbart zu den durch die Transkription des Audiosignals (AS) erhaltenen Textabschnitten (TP) in dem Dokument (DO) existieren, ausgebildet sind, wobei gegebenenfalls ein Unterbrechen des Wiedergebens des Audiosignals (AS) während des Wiedergebens der weiteren Textabschnitte (TP') bewirkbar ist.

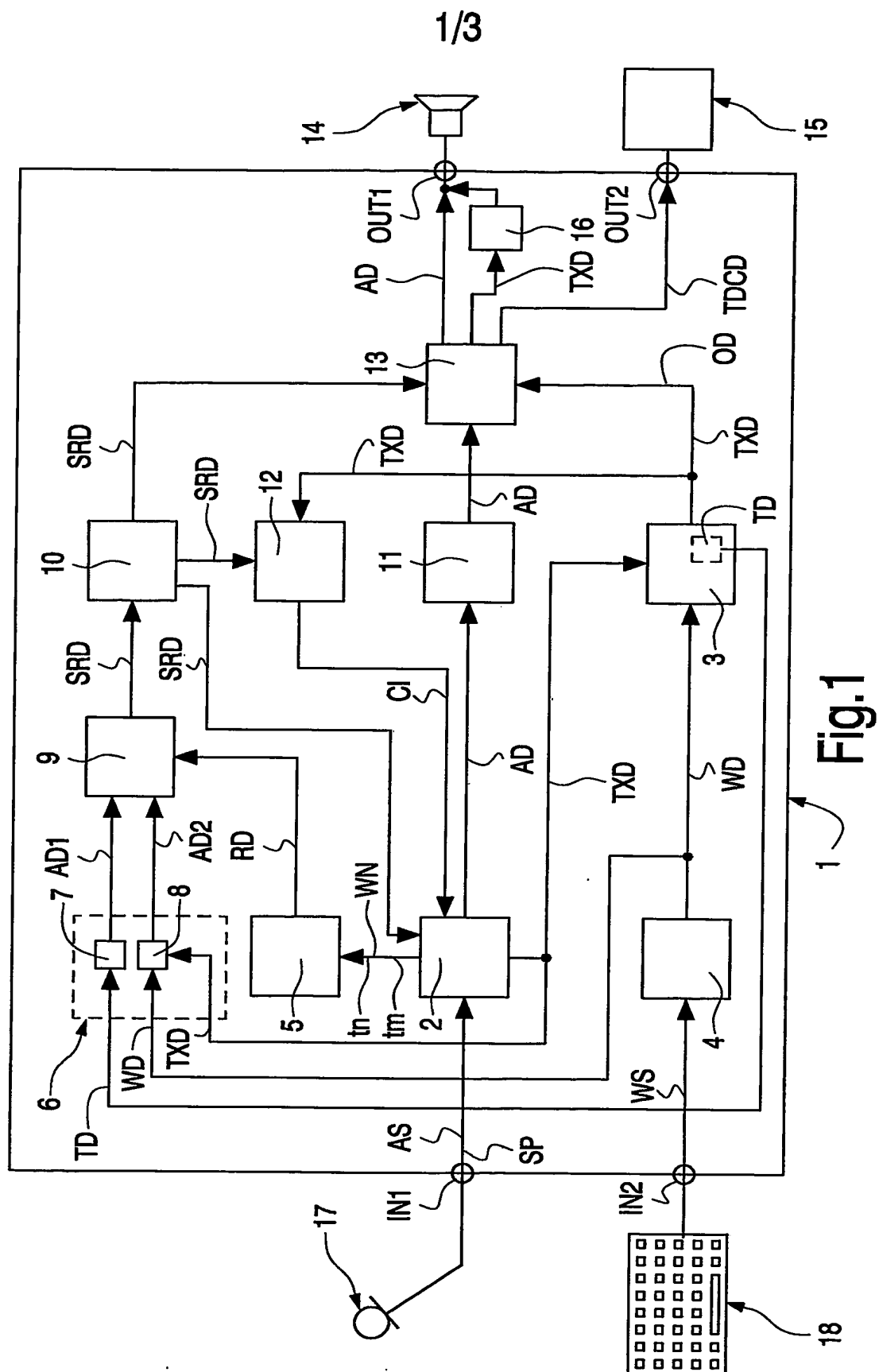
- 20 15. Computerprogrammprodukt,
das zum Transkribieren eines Audiosignals (AS) geeignet ist und das direkt in einen Speicher eines Computers geladen werden kann und Softwarecodeabschnitte umfasst, wobei mit dem Computer das Verfahren gemäß dem Anspruch 1 abgearbeitet werden kann, wenn das Computerprogrammprodukt auf dem
25 Computer abgearbeitet wird.

16. Computerprogrammprodukt nach Anspruch 15,
wobei das Computerprogrammprodukt auf einem computerlesbaren Medium gespeichert ist.

17. Computer mit einer Recheneinheit und einem internen Speicher, der das
30 Computerprogrammprodukt gemäß dem Anspruch 15 abarbeitet.

ZusammenfassungVerfahren und Einrichtung zum Transkribieren eines Audiosignals

- 5 Bei einem Verfahren zum Transkribieren eines Audiosignals (AS) enthaltend
Signalabschnitte (SP) in Text enthaltend Textabschnitte (TP) für ein Dokument (DO),
welches Dokument (DO) zum Wiedergeben von Informationen vorgesehen ist, welche
Informationen zumindest teilweise zu den durch die Transkription erhaltenen
Textabschnitten (TP) korrespondieren, ist vorgesehen, dass Signalabschnitte (SP) in
10 Textabschnitte (TP) transkribiert werden und Beziehungsdaten (RD), die zumindest eine
zeitliche Beziehung zwischen je einem Signalabschnitt (SP) und je mindestens einem
durch die Transkription erhaltenen Textabschnitt (TP) repräsentieren, erzeugt werden und
dass eine Struktur des Dokuments (DO) erkannt wird und dass die erkannte Struktur des
Dokuments (DO) in den Beziehungsdaten (RD) abgebildet wird.
- 15 (Figur 1)



2/3

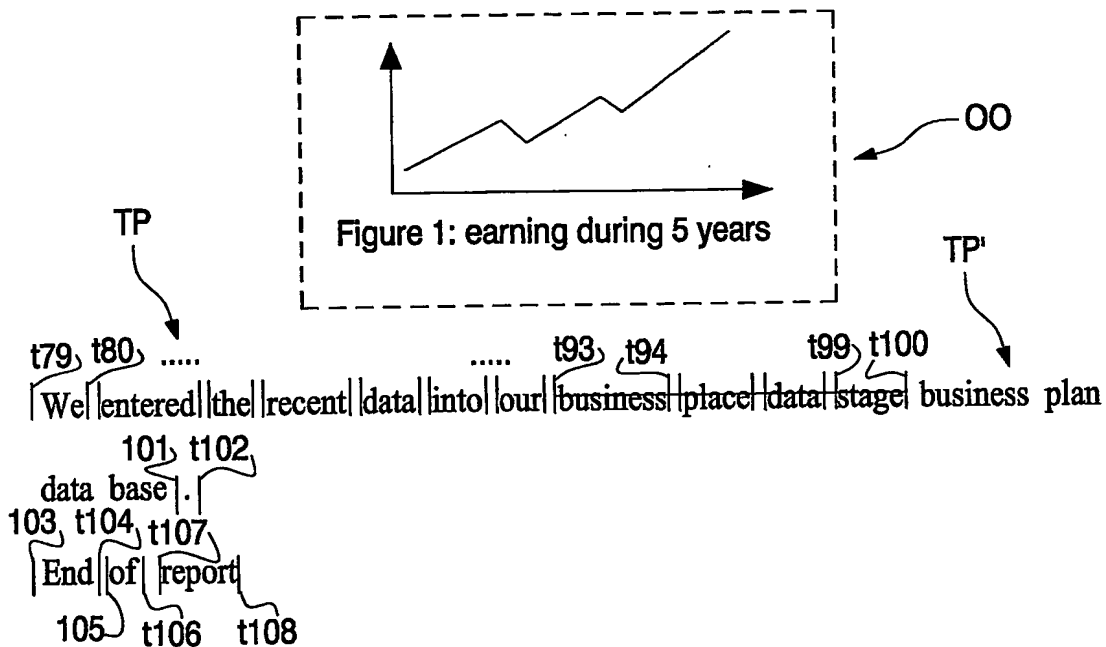
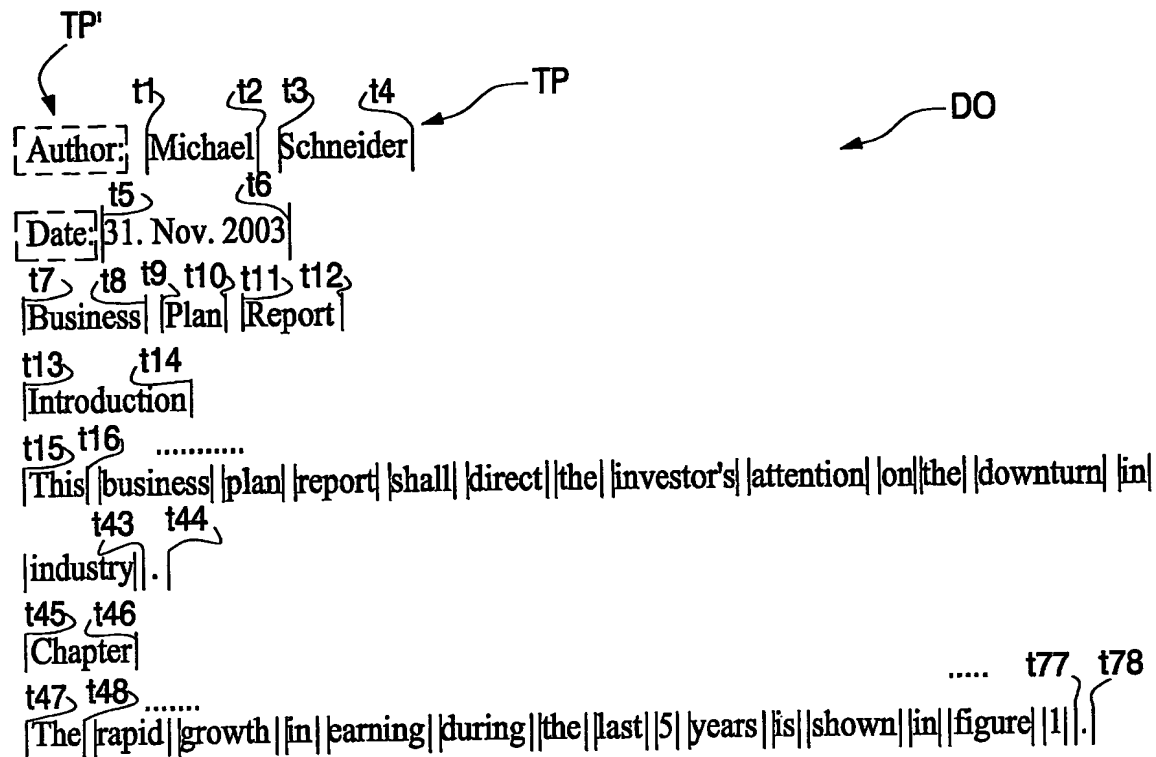


Fig.2

3/3

1	t1,	t2,	Text: Author	
...				
2	t3,	t4,	Text: Author	RD ↙
3	t5,	t6,	Text: Date	
4	t7,	t8,	Text: Report Heading	
...				
7	t11,	t12,	Text: Report Heading	
8	t13,	t14,	Text: Chapter Heading	
9	t15,	t16,	Text	
...				
22	t43,	t44,	Text	
23	t45,	t46,	Text: Chapter Heading	
24	t47,	t48,	Text	
...				
39	t77,	t78,	Text	
40	t78,	t79,	Object with Text but no audio	
41	t79,	t80,	Text	
...				
47	t91,	t92,	Text	
48	t93,	t94,	Text to skip	
...				
51	t99,	t100,	Text to skip	
52	t100,	t101,	Text without audio	
53	t101,	t102,	Text	
...				
56	t107,	t108,	Text	
C1				
C2				
C3				
C4				

Fig.3

PCT/IB2004/052529

